DETECTION-THEORETIC ANALYSIS OF WARPING ATTACKS IN SPREAD-SPECTRUM WATERMARKING

Alexia Briassouli and Pierre Moulin
{briassou,moulin}@ifp.uiuc.edu

University of Illinois
Beckman Inst., Coord. Sci. Lab & ECE Dept.
405 N. Mathews Ave., Urbana, IL 61801

Abstract. This paper studies the effects of desynchronization attacks such as delay and warping on the performance of blind spread-spectrum watermark detection systems. The host signal is modeled as a colored Gaussian signal. Evaluation of the optimal likelihood ratio test is often computationally expensive, so as a practical alternative, we propose a family of quadratic detectors and construct the detector and family of watermarks that maximize the deflection criterion. Experiments are carried out to verify the suitability of the deflection as a performance index. Substantial improvements over conventional watermark designs are demonstrated.

1. INTRODUCTION

Consider the problem of detecting a known watermark w originally embedded in a host signal s. The watermarked signal $x = s + w$ is subjected to attacks. The corrupted signal y is made available to the watermark detector, together with the reference watermark w. It is known that desynchronization attacks such as unknown delays and warping (time-varying delay) can disable empirically designed detectors [1]. A natural approach to combat such attacks is to formulate watermark detection as a composite hypothesis testing problem. This paper extends our recent work on white Gaussian hosts [2] and constructs a detector and a family of watermarks that are computationally tractable and satisfy optimality properties under warping attacks. The theory is however general enough to be applicable to a larger list of attacks.

2. MATHEMATICAL MODEL

For mathematical convenience, we assume that all signals are discrete-time and periodic with period equal to N. The host $s(n)$ is a Gaussian periodic process with full-rank covariance matrix $R_s = \{R_s(k,k') = E[s(k)s(k')], 1 \leq k, k' \leq N\}$. The special case of a white Gaussian host was considered in [2]. Unlike the setup in [2], here we consider a random watermark, because this increases the system security. The watermark considered is a zero-mean random signal, with covariance matrix R_w to be designed. The warping function $\theta(n),\ 1 \leq n \leq N$, is real-valued and slowly-varying. By $s(n-\theta(n))$ and $w(n-\theta(n))$ we denote the resampled versions of the underlying continuous-time warped signals.

The detection problem can be formulated as a composite hypothesis test [3]:

$$\begin{cases} H_0 & : y(n) = s(n), & 1 \leq n \leq N \\ H_1 & : y(n) = w(n-\theta(n)) + s(n-\theta(n)), & 1 \leq n \leq N. \end{cases} \tag{1}$$

We assume that the statistics of $s(n)$ are indistinguishable from those of $s(n-\theta(n))$, otherwise the host signal itself would serve as a synchronization signal. Hence, we study the hypothesis test

$$\begin{cases} H_0 & : y(n) = s(n) & ,\ 1 \leq n \leq N \\ H_1 & : y(n) = w(n-\theta(n)) + s(n) & ,\ 1 \leq n \leq N, \end{cases} \tag{2}$$

which serves as an approximation to the original detection problem.

3. ANALYSIS OF WARPING ATTACKS

3.1. Coherent Detector

If the warping function $\theta(n),\ 1 \leq n \leq N$ was known, we would have a coherent detection problem [3]. The likelihood ratio test (LRT) compares the linear statistic c_θ with a threshold η:

$$c_\theta = \sum_{k=1}^{N} \sum_{l=1}^{N} w(k-\theta(k)) R_s^{-1}(k,l) y(l) \ \underset{H_0}{\overset{H_1}{\gtrless}}\ \eta. \tag{3}$$

THIS WORK WAS SUPPORTED BY NSF GRANTS CCR 00-81268 AND CDA 96-24396.

Under H_0 and H_1, the statistic c_θ has means $m_0 = 0$ and

$$m_1 = \sum_{k=1}^{N}\sum_{l=1}^{N} w(k-\theta(k))R_s^{-1}(k,l)w(l-\theta(l)), \quad (4)$$

respectively, and variances

$$\sigma_0^2 = \sigma_1^2 = \sum_{k=1}^{N}\sum_{l=1}^{N} w(k-\theta(k))R_s^{-1}(k,l)w(l-\theta(l)). \quad (5)$$

For Bayesian detection under equal priors on H_0 and H_1, the threshold of the LRT is $\eta = \frac{1}{2}(m_0 + m_1)$, and the probability of error is $P_e = Q(\frac{1}{2}\sqrt{SNR})$, where $Q(u) = \int_u^\infty \phi(v)\,dv,\ \phi(u) = (2\pi)^{-1/2}\exp\{-u^2/2\}$ and

$$SNR \triangleq \frac{(m_1 - m_0)^2}{\sigma_1^2}. \quad (6)$$

3.2. Quadratic Noncoherent Detector

When θ is unknown, a suboptimal but often good approach to noncoherent detection consists of using a quadratic detection test [5] of the form

$$z = y^T K y \quad \underset{H_0}{\overset{H_1}{\gtrless}} \quad \eta \quad (7)$$

where K is an $N \times N$ symmetric matrix, and η is the threshold of the test.

3.2.1. Deflection Criterion

Assume the N-vector θ is random over $[0, N]^N$, with a distribution $\pi(\theta)$, and is independent of $s(n), 1 \le n \le N$. Computing the first two moments of Z in (7) under H_0 and H_1, we obtain

$$\begin{aligned} E_{w,\theta}[Z|H_0] &= Tr(R_sK), & (8)\\ E_{w,\theta}[Z|H_1] &= Tr(R_sK) + Tr(R_{w,\pi}K), & (9) \end{aligned}$$

where $R_{w,\pi}$ is an $N \times N$ watermark autocorrelation matrix with entries

$$\begin{aligned} R_{w,\pi}(k,k') &= E_{w,\theta}[w(k-\theta(k))w(k'-\theta(k'))] \\ &= E_\theta[R_w(k-\theta(k), k'-\theta(k'))] \\ &= \int\int R_w(k-\theta(k), k'-\theta(k')) \\ &\qquad \times p(\theta(k),\theta(k'))d\theta(k)d\theta(k') \quad (10) \end{aligned}$$

for $1 \le k, k' \le N$. Here $p(\theta(k),\theta(k'))$ denotes the bivariate probability density function of $\theta(k)$ and $\theta(k')$. [1] The set of all $R_{w,\pi}$ of the form (10) is the *feasible set* $\mathcal{R}_{w,\pi}$.

After some algebra, the variance of Z under H_0 is

$$Var[Z|H_0] = 2Tr(R_sKR_sK^T). \quad (11)$$

The threshold of the test (7) satisfies

$$Tr(R_sK) < \eta < Tr(R_{w,\pi}K) + Tr(R_sK). \quad (12)$$

We define the *deflection criterion* (also called generalized SNR) for quadratic detection as [5]

$$\begin{aligned} d^2 &= \frac{(E_{w,\theta}[Z|H_1] - E_{w,\theta}[Z|H_0])^2}{Var[Z|H_0]} \\ &= \frac{(Tr(R_{w,\pi}K))^2}{2Tr(R_sKR_sK^T)}. \quad (13) \end{aligned}$$

This criterion would determine the probability of error of the test (7) if the distributions of Z under H_0 and H_1 were Gaussian. In problems such as ours, it only serves as a tractable measure of separability of the two distributions: higher values of the deflection are expected to lead to lower error probabilities. It can be shown that d^2 is maximized by

$$K = \alpha R_s^{-1} R_{w,\pi} R_s^{-1} \quad (14)$$

where α is an arbitrary nonzero constant. The maximum value of the deflection for the optimal kernel K is:

$$d^2 = \frac{1}{2} Tr(R_{w,\pi}R_s^{-1}R_{w,\pi}R_s^{-1}). \quad (15)$$

Remark. The mean-square average of c_θ in (3) is:

$$E_{w,\theta}[c_\theta^2] = \sum_{k=1}^{N}\sum_{l=1}^{N} y(k)M(k,l)y(l) = y^T M y \quad (16)$$

where $M = R_s^{-1}R_{w,\pi}R_s^{-1}$. Comparing with (14), we conclude that (16) is an optimal quadratic decision statistic.

3.2.2. Optimal Watermark Design

The use of d^2 as a performance criterion for quadratic detection also suggests its use as a criterion for watermark design. The criterion d^2 depends on the statistics of the watermark only via its correlation matrix R_w.

The watermark should be imperceptible, so we constrain its average energy per sample:

$$\frac{1}{N} Tr(R_{w,\pi}) \le \sigma_w^2. \quad (17)$$

The maximization of (15) subject to the constraint of (17) and the constraint $R_{w,\pi} \in \mathcal{R}_{w,\pi}$ (as defined below (10)) must generally be done numerically. However, a useful upper bound on d^2 can be derived. Let $\Lambda_s = diag\{\lambda_s(k),\ 1 \le k \le N\}$ and $\Lambda_{w,\pi} = diag\{\lambda_{w,\pi}(k),\ 1 \le k \le N\}$ be diagonal matrices made of the eigenvalues of R_s and $R_{w,\pi}$

[1] If moreover $\theta(k)$ is a periodic stationary stochastic process with uniform marginal distributions, then $p(\theta(k),\theta(k'))$ depends only on $k - k'$ and on $\theta(k) - \theta(k')$, and it can be verified that $R_{w,\pi}$ is Toeplitz.

respectively. The elements of Λ_s and $\Lambda_{w,\pi}$ are nonnegative and are arranged in the same order. Then, from (15) we have [4]:

$$d^2 \leq \frac{1}{2}\sum_{k=1}^{N}\left(\frac{\lambda_{w,\pi}(k)}{\lambda_s(k)}\right)^2. \qquad (18)$$

The upper bound (18) is achieved when both $R_{w,\pi}$ and R_s^{-1} have the same eigenvector matrix U. This fact will play an essential role in the analysis. Maximization of the right side of (18) subject to the constraint

$$\frac{1}{N}Tr(R_{w,\pi}) = \frac{1}{N}\sum_{k=1}^{N}\lambda_{w,\pi}(k) \leq \sigma_w^2, \qquad (19)$$

leads to the following upper bound on the deflection:

$$d^2 \leq d_{ub}^2 = \frac{N^2\sigma_w^4}{2\lambda_{s,min}^2}. \qquad (20)$$

where $\lambda_{s,min}$ is the smallest eigenvalue of the host. This upper bound is attained by some $R_{w,\pi}^*$ if and only if $R_{w,\pi}^* = U\Lambda_{w,\pi}^*U^H$, where ^{H} denotes Hermitian transpose, the eigenvector matrix of $R_{w,\pi}^*$ is the same as that of R_s, and

$$\Lambda_{w,\pi}^* = diag(0,...,0,N\sigma_w^2,0,...,0). \qquad (21)$$

In other words, all the available power is assigned to one single eigenvalue corresponding to $\lambda_{s,min}$. The upper bound on the deflection can be achieved only if $R_{w,\pi}^* \in \mathcal{R}_{w,\pi}$. However, in general, it is not always possible to construct a feasible $R_{w,\pi}$ that achieves this upper bound. A case of interest, shown in Example 2, is when R_s is circulant Toeplitz and a feasible $R_{w,\pi}^*$ can be generated.

Once a feasible $R_{w,\pi}^*$ is found, we still need to find a corresponding correlation matrix R_w^*. Equation (10) defines a linear mapping $R_{w,\pi}^* = \mathcal{L}_\pi R_w^*$. Any $R_w \in \mathcal{L}_\pi^{-1}(R_{w,\pi}^*)$ is therefore optimal.

Example 1. A case of practical interest, where R_w^* can be found easily from $R_{w,\pi}^*$, occurs when the attack is a simple delay that is uniformly distributed, and R_s and $R_{w,\pi}^*$ are circulant Toeplitz. In this case, the choice $R_w^* = R_{w,\pi}^*$ satisfies (10). For example, the watermark $w(n)$ could be generated as a Gaussian signal with zero mean and correlation matrix $R_w^* = R_{w,\pi}^*$.

Example 2. Consider a warping attack satisfying the conditions of footnote 1 and assume again that R_s and $R_{w,\pi}^*$ are circulant Toeplitz. Then $R_{w,\pi}$ is circulant Toeplitz. We seek R_w that satisfies:

$$\begin{aligned} R_{w,\pi}^*(k,k') &= \frac{1}{N}\int\int R_w(k-\theta(k),k'-\theta(k')) \\ &\qquad \times p_{k-k'}(\theta(k)-\theta(k'))d\theta(k)d\theta(k') \\ &= \int R_w(k-k'-\Delta)p_{k-k'}(\Delta)d\Delta \qquad (22) \end{aligned}$$

where $p_{k-k'}$ is the distribution of $\theta(k)-\theta(k')$. A circulant Toeplitz solution R_w is guaranteed to exist.

Detectability/Security Tradeoff. In an actual watermarking application, it may be desirable to spread the watermark power over several eigenvectors to increase the system's security. For instance, let the power be equally distributed among the weakest L eigenvectors of the host, i.e., the power allocated to each channel is $\lambda_w(k) = N\sigma_w^2/L$, where $1 \leq k \leq L$. In that case we have

$$d^2 = \frac{1}{2}\sum_{k=1}^{L}\left(\frac{\lambda_{w,\pi}(k)}{\lambda_s(k)}\right)^2 \leq \frac{1}{L}\frac{N^2\sigma_w^4}{2\lambda_{s,min}^2} = \frac{1}{L}d_{ub}^2 \qquad (23)$$

If $\lambda_s(k),\ 1 \leq k \leq L$, are all equal, (23) is achieved with equality. The deflection decreases when the available watermark power is distributed among many channels of the host. Thus, there is a trade-off between detectability and security.

Example 3. Consider a uniformly distributed delay attack and a periodic and stationary AR(1) host s. Then U is the DFT matrix, and $R_{w,\pi} = U\Lambda_{w,\pi}U^H$ is circulant Toeplitz. When the watermark power is allocated entirely to $k = N/2$, the weakest component of the host, we have $\Lambda_{w,\pi} = \Lambda_{w,\pi}^*$. Here, $R_{w,\pi}^* = U\Lambda_{w,\pi}^*U^H \in \mathcal{R}_{w,\pi}$, so the upper bound d_{ub}^2 on the deflection is achieved. Referring to Example 1, we can choose $R_w^* = R_{w,\pi}^*$. The resulting optimal watermark is a sinusoid with frequency π. The watermark power could also be assigned to two eigenvalues to increase the system security and also ensure that the resulting watermark is real-valued. Then, if $\lambda_w(k) = \lambda_w(N-k)$ for $k \neq 0, k \neq N/2$, $w(n)$ becomes a real sinusoidal watermark at frequency $2\pi k/N$, and $d^2 = \frac{1}{2}d_{ub}^2$. Similarly, the watermark power can be equally distributed among an even number of the eigenvalues to further increase the system security, but at the cost of a lower d^2. The resulting watermarks will be sums of sinusoids at the frequencies chosen.

4. NUMERICAL RESULTS

We performed several experiments and evaluated the performance of our design numerically, in terms of the deflection criterion and the probability of error (which as mentioned earlier is not a function of the deflection criterion) under different choices of the kernel and different choices of the watermark. Monte-Carlo runs (averaging over all random variables) were used to determine empirical probabilities of error. In order to draw reliable and generally applicable conclusions, we considered watermarks of varying strength, parameterized by the ratio of the maximum host power to the maximum watermark power. The ratio σ_s^2/σ_w^2 was in the range of 8 to 20 dB. We considered a length $N = 400$ periodic Gaussian AR(1) host with zero mean, unit variance and $\rho = 0.98$ displayed in Fig. 1. Assume the warping attack is a periodic AR(1) process with mean zero, variance

$\sigma_e^2 = (\rho + 1)\epsilon^2/2$ (where $\epsilon = 0.04$, see [2]) and $\rho = 0.98$ (see example in Fig. 1). The optimal watermark under this setup is sinusoidal with frequency π, as all the power is allocated to $\lambda_w(N/2)$.

Experiments were also conducted to compare the performance of the proposed scheme with the optimal kernel K of (14) and with $K = R_w$ (which would be optimal for a white host [2]). As Fig. 2(a) shows, the value of the deflection indeed decreases when the suboptimal kernel is used, suggesting that the distributions under H_0 and H_1 are less separated than they are when the optimal kernel is used. Consequently, the error probability is expected to increase. Indeed, Fig. 2(b) shows that use of the optimal kernel leads to significantly lower error probabilities (ranging from 0.005 to 0.2 instead of 0.01 to 0.45). The error probability for the coherent detector (which knows the warping function and serves as an oracle) is also shown in Fig.2(b). It is of course lower than for the noncoherent case, but the gap is small when the host-to-watermark power ratio σ_s^2/σ_w^2 is large.

Using the optimal kernel, experiments were also conducted to compare the performance of an optimal watermark, designed following Section 3.2.2, with one generated from a suboptimal covariance R_w with the same structure as R_s. The watermark is statistically similar to the host signal. Fig. 3 shows that the deflection increases and the probability of error decreases when an optimal watermark is hidden.

5. REFERENCES

[1] D. Kirovski and H. Malvar, “Robust Covert Communication over a Public Audio Channel using Spread Spectrum,” *Proc. Information Hiding Workshop*, Pittsburgh, PA, 2001.

[2] P. Moulin, A. Briassouli and H. Malvar, “Detection-Theoretic Analysis of Desynchronization Attacks in Watermarking,” *DSP 2002, 14^{th} Int. Conf. on DSP*, Santorini, Greece, July 2002.

[3] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd Ed., Springer-Verlag, 1994.

[4] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge, 1996.

[5] C. R. Baker, “Optimum Quadratic Detection of a Random Vector in Gaussian Noise,” *IEEE Trans. on Comm.*, Vol. 14, No. 6, pp. 802—805, 1966.

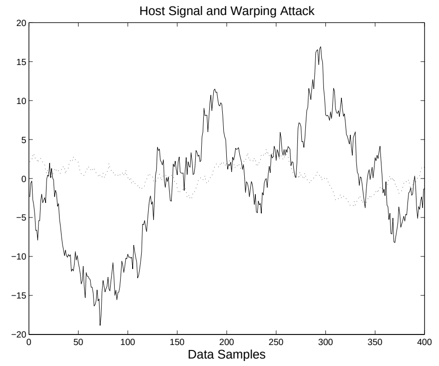
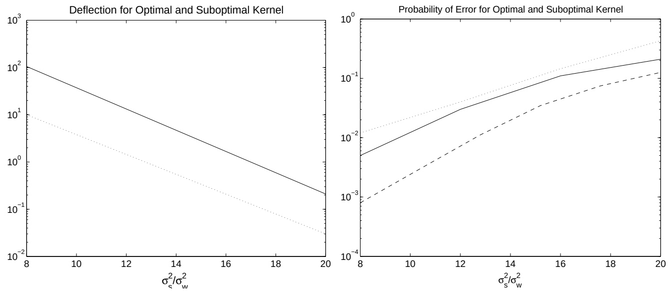
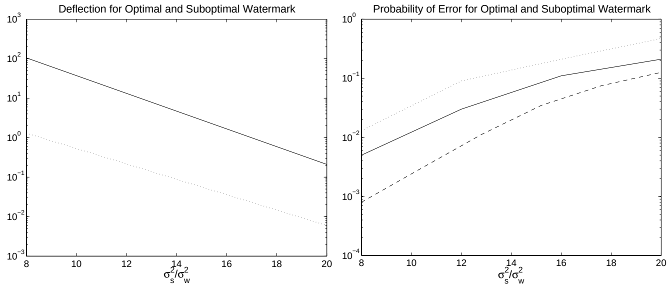


Fig. 1. Gaussian AR(1) host with $\rho = 0.98$ (solid line) and AR(1) warping attack with $\rho = 0.98$ (dotted line).



(a) (b)

Fig. 2. The effect of different kernels K on the system performance for optimally designed watermarks of different strength. (a) Deflection (b) Probability of error for $K_{opt} = R_s^{-1} R_w R_s^{-1}$ (solid line), $K = R_w$ (dotted line) and coherent detection (dashed line).



(a) (b)

Fig. 3. The effect of suboptimal watermarks of different strength on the system performance for optimally designed kernels K. (a) Deflection (b) Probability of error for an optimal R_w (solid line), for R_w with the same structure as R_s (dotted line) and coherent detection (dashed line).